SMALL RNA (sRNA) REGULATES EXTREME TEMPERATURE ADAPTATION OF Geobacillus thermoleovorans ISOLATED FROM A MALAYSIAN HOT SPRING

TAN HOCK SIEW

UNIVERSITI SAINS MALYSIA

2011

SMALL RNA (sRNA) REGULATES EXTREME TEMPERATURE ADAPTATION OF *Geobacillus thermoleovorans* ISOLATED FROM A MALAYSIAN HOT SPRING

by

TAN HOCK SIEW

Thesis submitted in fulfillment of the requirements
for the degree of
Master of Science

February 2011

RNA KECIL (sRNA) MENGAWALATUR ADAPTASI TERHADAP SUHU LAMPAU *Geobacillus thermoleovorans* YANG DIPENCIL DARI MATA AIR PANAS MALAYSIA

oleh

TAN HOCK SIEW

Tesis yang diserahkan untuk
memenuhi keperluan bagi
Ijazah Sarjana Sains

Febuari 2011

ACKNOWLEDGEMENTS

I would like to thank my main supervisor Prof Maqsudul Alam for his guidance and advice throughout my Masters project. He had provided me with the resources to advance in this field of research. I would also like to thank my co-supervisors Prof Nazalan Najimudin and Prof Darah Ibrahim for their critical reading and advice. My thanks also go to Dr. Jonathan Livny who has helped me to set up the data analysis pipeline.

It is always encouraging and exciting to work in a lab that has supportive lab mates. I would like to thank them for their help rendered throughout my work in the lab. They are there for you when you need them and this made it a conducive environment for scientific research. I would like to also thank Dr. Jennifer Saito for her helpful suggestions for my work.

My special thanks to my mum and fiancée, Erin for being ever understanding of the nature of my project. Your unconditional love and support helped me throughout the trying times of this project.

TABLE OF CONTENTS

CHAPTER 2 – LITERATURE REVIEW

CHAPTER 3 – MATERIALS AND METHODS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

APS	Ammonium Persulfate
AS	Antisense
BLAST	Basic Local Alignment Search Tool
CRISPR	Clustered regularly interspaced palindromic repeats
DEPC	Diethylpyrocarbonate
DMSO	Dimethyl sulfoxide
dNTP	Deoxynucleotides triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
GYT	Glucose yeast tryptone media
HSR	Heat shock response
HTR	High temperature response
IGR	Intergenic region
$K_2 HPO_4$	Dipotassium phosphate
miRNA	Micro RNA
$MgCl_2$	Magnesium chloride
NaCl	Sodium chloride
ncRNA	Non-coding RNA
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
rRNA	Ribosomal RNA

RNAP	RNA polymerase
siRNA	Small interfering RNA
TBE	Tris borate EDTA Buffer
TEMED	Tetramethylethylenediamine
tRNA	Transfer RNA
UTR	Untranslated region

RNA kecil (sRNA) mengawalatur adaptasi terhadap suhu lampau *Geobacillus thermoleovorans* yang dipencil dari mata air panas Malaysia

Abstrak

RNA kecil (sRNA) telah ditunjukkan memainkan peranan penting dalam pengawalan gen dalam eukariot (juga dikenali sebagai microRNA) dan prokariot. Dalam bakteria, sRNA mengawalatur pembentukan biofilem, penderian korum, kevirulenan dan gerak balas stres persekitaran. Antara pelbagai gerak balas stres, tindak balas kejutan suhu telah banyak dikaji dalam bakteria patogen dan extremofil. sRNA telah dipostulatkan terlibat dalam adaptasi bakteria terhadap suhu lampau melalui rangkaian pengawalaturan stres sejagat. Fokus kajian ini adalah terhadap sRNA yang terlibat dalam perkembangan bakteria thermofil, *Geobacillus thermoleovorans* CCB_US3_UF5 pada suhu 60°C dan 70°C. sRNA daripada pertumbuhan pada suhu lampau ini telah dilakukan proses transcripsi berbalik kepada cDNA dan dijujuk. Data penjujukan mengenalpasti 83 sRNA putatif yang diklasifikasikan sebagai antisense, kawasan di antara gen, kawasan tidak bertranslasi atau bukan pengkodan. Daripada jumlah ini, 44 calon sRNA adalah spesifik terhadap pertumbuhan pada suhu tinggi. Kajian ini mendemostrasikan satu alir kerja lengkap analisis sRNA untuk bakteria thermofil dengan mengabungkan satu protokol pengklonan sRNA dan salur pipa menganalisis data penjujukan truput tinggi. Secara keseluruhannya, keputusan ini dapat digunakan untuk kajian masa hadapan untuk memahami dengan lebih terperinci pengawalaturan adaptasi bakteria termofil pada suhu lampau oleh sRNA.

Small RNA (sRNA) regulates extreme temperature adaptation of *Geobacillus thermoleovorans* isolated from a Malaysian hot spring

Abstract

Small RNA (sRNA) has been shown to play important gene regulatory roles in both eukaryotes (commonly known as microRNA) and prokaryotes. In bacteria, sRNAs regulate biofilm formation, quorum sensing, virulence, and environmental stress response. Of the various stress responses, the heat shock response has been extensively studied in many pathogenic bacteria and extremophiles. The focus of this study is to identify the sRNAs that may be involved in the adaptation to 60°C and 70°C of a thermophilic bacterium, *Geobacillus thermoleovorans* CCB_US3_UF5. The sRNAs from these extreme high temperature growths were reverse transcribed to cDNA and sequenced. Sequencing data identified 83 putative sRNAs classified as antisense, intergenic region, untranslated region, or non-coding. Out of this total, 44 sRNA candidates were specific to growth at elevated temperature. This study demonstrated a complete sRNA analysis workflow for a thermophilic bacteria by combining a sRNA cloning protocol and high-throughput sequencing data analysis pipeline. Collectively, these results can be used for future studies to better understand the detailed sRNA regulation of extreme temperature adaptation in thermophilic bacteria.

Why small RNA?

I have always been fascinated by the ability of bacteria to survive in any known environment in our planet. How do the bacteria know what type of environment they are in? How do they cope with a stressful environment? How do they regulate their cell machinery in an efficient and coordinated manner for survival? These are some of the fundamental questions that have intrigued me for a long time. Over the years, many studies have elucidated some of these questions. Proteins had been extensively studied for their role in regulatory networks probably because they are the end product in a cell, signifying its functional role at the particular environment. In the last decade, RNA and DNA have been receiving more attention and they began to connect the missing link in the bacterial regulatory network.

Despite being short in length and usually not coding for any proteins, the regulatory capability of small RNAs (sRNAs) enabled bacteria to respond to changing environmental signals rapidly. The underlying mechanisms of gene regulation by these small RNA drive my interest to study these small RNA.

CHAPTER 1

INTRODUCTION

1.1 Extremophiles

Microorganisms can be found in almost every part of the world. They are very versatile and are able to adapt under many different environmental conditions. Among many different types of microbes, one group can live in harsh environments. They are known as the extremophiles. These extreme-loving microbes can live in very high or low temperature, extreme pH, salinity and some may even tolerate high dosage of gamma radiation.

In Malaysia, hot springs are one of the most prominent features of extreme environments. The Ulu Slim hot spring was reported to be one of the hottest hot springs in Malaysia. Two thermophilic or heat-loving bacteria have been isolated from this hot spring and one of them is *Geobacillus thermoleovorans* CCB_US3_UF5.

In order to cope with heat stimuli, there are two major responses in bacteria depending upon when the heat stress is applied. The heat shock response (HSR) occurs when there is a sudden increase in temperature. In contrast the high temperature response (HTR) that occurs when the bacteria is already growing at a constant elevated temperature. High temperature response genes have to be expressed if any of these conditions occur. The expression of these genes is highly regulated and involves regulatory proteins and RNAs. One of the RNA regulators in bacteria is small RNA.

1.2 Small RNA

Small RNA or sRNA is a type of regulatory RNA. The role of sRNA is to regulate physiological changes in bacteria in response to environmental stimuli. It is generally 20-550 nucleotides (nt) in length. It is single-stranded and able to form a stem loop that confers greater structural stability. sRNAs exert their effect at the post-transcriptional level of gene expression. This causes either activation or deactivation of the target mRNA expression.

sRNA regulates its target via two major mechanisms. One is by base-pairing with its target mRNA while the other is by direct binding to regulatory proteins. Among these two mechanisms, the base-pairing mechanism has been extensively studied and the sRNAs that fall into this category can be further divided into two classes. The basis of the classification is dependent on the location where the sRNA is encoded in the genome. One is known as the *trans*-encoded sRNA while the other is known as *cis*-encoded sRNA. Further discussions on these sRNAs are covered in the literature review.

1.3 Rationale of the study

HSR is triggered upon sudden temperature elevation. In contrast, HTR involves survival and growth at a higher temperature that the cells have accustomed to over a few generations. Acute phase and chronic phase are part of the adaptation process. In the initial acute phase, rapid responses are needed to help the cell survive in their new environment. In the chronic phase, the cells have survived and begin to grow exponentially in the new environment with possibly a new growth rate. Interestingly the group of genes that are involved in the acute phase are different from those involved in the chronic phase (Gunasekera *et al.*, 2008).

Regardless of whether the HSR or HTR system is used, sRNAs play a vital role in both. Almost every environmental stress factor involves a specific pathway that may require sRNAs to regulate their genes. The only sRNA that has been reported so far that responds to a temperature shift is DsrA in *E. coli* (Narberhaus and Vogel, 2009). However, DsrA regulates its target during low temperature (Sledjeski *et al.*, 1996). None have been reported at high temperature. In addition, no sRNAs have been reported from thermophilic bacteria that are involved in HTR or HSR.

I am intrigued to know which sRNAs are specifically induced during high temperature growth (HTR) *in vitro* by using a thermophilic bacterium, *G. thermoleovorans* CCB_US3_UF5 as our model organism. To achieve this goal, I will be cloning and sequencing the sRNAs from this bacterium.

One of the major challenges in cloning sRNA from bacteria is the presence of tRNA and 5S rRNA in the size region of sRNA (20-550 nt). tRNA have an average length of 80 nt while 5S rRNA is usually 125 nt. Since both tRNA and 5S rRNA are in abundance and may reduce the probability of obtaining sRNA sequences during high throughput sequencing, they have to be depleted from the sRNA that will be cloned. Recently, a method for tRNA and 5S rRNA depletion has been performed for sRNA samples from *Vibrio cholera (*Liu *et al., 2009)*. This method will be adapted for the use in cloning sRNA from *Geobacillus thermoleovorans* CCB_US3_UF5.

1.4 Objectives

1.4.1 To establish an sRNA cloning system that enriches sRNA in bacteria.

1.4.2 To identify sRNAs expressed during extreme temperature adaptation.

CHAPTER 2

LITERATURE REVIEW

2.1 Extremophiles

Extreme environments are often considered to be either physically or geochemically extreme for the survival of most life forms on earth. pH and salinity are considered as geochemical extremities while radiation, pressure and temperature are classified as physical extremities (van den Burg, 2003). Organisms that survive and thrive in these harsh conditions are known as extremophiles.

These extremophiles have been found in seven most extreme conditions in this planet. Table 2.1 summarizes the different types of extremophiles. Many extremophiles are tolerant of more than one geochemical or physical extremity. For example, *Thermoplasma acidophilum* is both thermophilic and acidophilic since it has the ability to survive at high temperature (59°C) and acidic conditions (pH 2) (Ruepp *et al.*, 2000).

Table 2.1 Classes of extremophiles and their extreme habitats (Pikuta *et al.*, 2007).

Geochemical/ Physical Parameter	Type of extremophily	Habitat	Description
Temperature	Thermophile	Hydrothermal vent, hot springs	Grows at 60°C and above
	Psychrophile	Polar ice sheets, permafrost, deep sea floor, snow cap, sea ice	Inability to grow at 15°C and above
pH	Acidophile	Hydrothermal springs, Acidic mine drainage	Grows at pH 3 or below
	Alkalidophile	Can be found in acidic or neutral soil (interesting characteristic compared to other extremophiles), soda lake	Requires at least pH 8 or above to grow (up to pH 13)
Pressure	Barophile	Hydrothermal vents, Deep ocean floors, deep oil reserves	Grows optimally growth at pressures more than 40 MPa
Salinity	Halophile	Hypersaline lake, salt springs	Does not depend on high salt concentration to survive but can tolerate up to 15% salinity
Radiation	Radiotolerant	Food sterilization process, hot spring	Able to withstand high doses of radioactive radiation. Highest: 800 kGy in *Deinococcus radiodurans* Lowest: 0.82 kGy in *Lactobacillus sakei*

In Malaysia, we can find hot geological environments such as hot springs. For example, thermophilic bacteria have been isolated from local hot springs (Akanbi *et al.*, 2010). One of the thermophiles that was isolated from Malaysian soil is *Geobacillus zalihae* which can thrive at 60°C (Abd Rahman *et al.*, 2007).

2.2 *Geobacillus* spp.

Geobacillus spp. are rod shaped, spore-forming, thermophilic bacilli. The genus *Geobacillus* was proposed in 2001 (Nazina *et al.*, 2001) by reclassifying thermophilic bacilli in phylogenetic group 5 of the *Bacillus* genus (Ash *et al.*, 1991).

The name *Geobacillus* carries the meaning earth or soil *Bacillus*. Members of this genus are listed in Table 2.2.

Table 2.2 Members of *Geobacillus* genus

Geobacillus stearothermophilus	(Logan and Berkeley, 1984; Priest *et al.*, 1988)
Geobacillus thermocatenulatus	(Nazina *et al.*, 2001; Nazina *et al.*, 2004)
Geobacillus vulcani	(Caccamo *et al.*, 2000; Nazina *et al.*, 2004)
Geobacillus gargensis	(Nazina *et al.*, 2004)
Geobacillus kaustrophilus	(Priest *et al.*, 1988; White *et al.*, 1993)
Geobacillus thermoglucosidasius	(Priest *et al.*, 1988; White *et al.*, 1993)
Geobacillus thermodenitrificans	(Manachini *et al.*, 2000)
Geobacillus subterranues	(Nazina *et al.*, 2001)
Geobacillus uzenensis	(Nazina *et al.*, 2001)
Geobacillus caldoxylosilyticus	(Fortina *et al.*, 2001)
Geobacillus toebii	(Sung *et al.*, 2002)
Geobacillus jurassicius	(Nazina *et al.*, 2005)
Geobacillus zalihae	(Abd Rahman *et al.*, 2007)
Geobacillus thermoleovorans	(Nazina *et al.*, 2001)

The *Geobacillus* genus has been known for their special physiological characteristics that have biotechnological importance. One of the widely studied properties of this genus is their thermostable enzymes such as proteases, amylases, lipases and pullanases (McMullan *et al.*, 2004). In a recent publication, *Geobacillus* spp. (M-7) was reported to produce volatile antibiotics capable of inhibiting test organisms including a human lung pathogen (*Aspergillus fumigates*) and other plant pathogens (*Geotrichum candidum, Botrytis cinerea, Ventricillum dahlia*) (Ren *et al.*, 2010). Its potential application of exopolysaccharides for industrial usage have also been discussed (Schiano Moriello *et al.*, 2003). Two strains of *G. thermoleovorans* have been found to produce bacteriocins that cause lytic activity on other *G.*

thermoleovorans strains and human pathogens including *Salmonella typhimurium* (Novotny and Perry, 1992). Different strains of *G. thermoleovorans* have been shown to degrade different types of compounds such as naphthalene, phenol, cresol, xylan, and olive oil (Deflaun *et al.*, 2007).

Geobacillus are able to grow aerobically in temperatures ranging from 35°C to 78°C (Nazina *et al.*, 2001) and are widely found in different geographical areas on Earth (White *et al.*, 1993). This group of thermophilic bacilli had been isolated from hot springs (Markossian *et al.*, 2000) and deep-sea hydrothermal vents (Maugeri *et al.*, 2002). They can also be found in high-temperature petroleum reservoirs (Nazina *et al.*, 2001) and temperate soil (Marchant *et al.*, 2002a). There are also reports of their isolation from man-made environments such as hot water pipelines, waste treatment plants and coal burning plants (Maugeri *et al.*, 2001; Obojska *et al.*, 2002). Although *Geobacillus* have been mostly found in hot environments, they have been isolated from cool environments as well such as soil samples from northern Ireland that have never reached more than 40°C (Marchant *et al.*, 2002a; Marchant *et al.*, 2002b).

2.3 Heat stress response in bacteria

Thermophiles that thrive in hot environments must constantly monitor the fluctuating environmental temperature for survival. This is because exposure to high temperature will cause many cellular stresses such as oxidative stress, protein denaturation, and may lead to lethal effects in the cell.

Apart from extremely high temperature, there are many biological processes in cells that are induced and controlled by temperature shift, including the cold shock response and expression of virulence genes (Narberhaus *et al.*, 2006). Microbial cells

have developed a wide array of thermosensors ranging from DNA to RNA to protein (Klinkert and Narberhaus, 2009).

There are two major heat stress responses in bacteria, depending on the induction mechanism (Table 2.3): the heat shock response (HSR) and the high temperature response (HTR) (Schumann, 2007). A classical example of HSR is the transfer of *E. coli* growing at 30°C to 42°C and *Bacillus subtilis* from 37°C to 48°C. HSR genes are upregulated by many folds upon temperature shift (Schumann, 2007). However, the expression of HSR genes is transient. There is a negative feedback loop (about 10 minutes after induction) in the expression of HSR genes ensuring that these genes are expressed only at two- to three-fold higher than the basal level of uninduced cell. HTR genes, on the contrary, are constitutively expressed as long as the cells are exposed to the elevated temperature.

Table 2.3: Summary of heat stress responses (adapted from Schumann, 2007).

Heat shock response	**High temperature response**
1. Response to sudden temperature increase	Response to absolute temperature
2. Temporary expression (Negative feedback loop)	Constitutive expression (Absent of feedback loop)
3. Expression of heat shock genes	Expression of high temperature genes

2.3.1 Heat shock response

Heat shock response has been extensively studied in *E. coli* (Guisbert *et al.*, 2008). This response induces the expression of a set of proteins called the heat shock proteins. The alternative sigma factor, σ^{32} drives the HSR gene expression rapidly to achieve a new steady-state level characteristic of the new temperature (Straus *et al.*, 1987). The majority of heat shock proteins consist of chaperones and proteases which help in the quality control of the proteins produced under stress (Gottesman *et*

al., 1997). Chaperones help to refold incorrectly folded proteins, while proteases degrade proteins that are severely misfolded or damaged (Bukau, 1993; Tomoyasu *et al.*, 2001).

Upon exposure to sudden temperature change, the HSR transcripts rapidly increase to its peak and are further maintained for another 10 to 15 minutes. The transcripts level then decreases to a new steady-state level for the new temperature, even though the cells are still subjected to continuous heat stress (Tomoyasu *et al.*, 1998). The "titration" model was proposed to explain this phenomenon. According to this model, since unfolded or misfolded proteins are the substrate for chaperones and proteases, the level of HSR is regulated based on the amount of these damaged proteins (Straus *et al.*, 1990; Tomoyasu *et al.*, 1998; Guisbert *et al.*, 2004). The downregulation of HSR after a short period of induction is crucial for the growth of the cells. Severe growth inhibition was observed when HSR genes are overexpressed (Herman *et al.*, 1995; Guisbert *et al.*, 2004).

In hyperthermophilic bacteria *Thermotoga maritima*, heat shock response was evident when the cells were subjected to temperature shift from 80°C to 90°C. It has a similar response to mesophilic bacteria although the heat stress regulatory strategies may be different (Pysz *et al.*, 2004).

2.3.2 High temperature response

The adaptation process to environmental stresses involves two main stages: the acute phase and chronic stage. During the acute phase, an immediate response is required for survival. In the chronic phase, the focus is shifted to support exponential growth once the cells have adjusted to the new environment (Gunasekera *et al.*, 2008).

In contrast to HSR that plays its role during the acute phase (sudden exposure), HTR deals with the chronic phase (continuous exposure) of the heat stress response. HSR is commonly found in mammalian pathogenic bacteria. The host temperature is controlled at about 37°C, which is higher than the environmental temperature but not lethal to mesophilic pathogens. Upon entry into the host, pathogenic bacteria need a signal (in this case the temperature of the host body) to indicate successful invasion of the warm-blooded mammal and virulence genes can then be expressed (Konkel and Tilly, 2000).

In a recent study of continuous osmotic and heat stress in *E. coli* (Gunasekera *et al.*, 2008), it was found that there is a poor correlation between the genes that are expressed during HSR and HTR. Genes that were classically found to be highly expressed during HSR (*dnaK, dnaJ, grpE, groEL, groES, clpB*) (Richmond *et al.*, 1999) were found to be only moderately expressed (~1.6 fold) during continuous heat stress. Instead, the genes that are involved in sulfur metabolism were highly upregulated (at least 4-fold) due to HTR. This study provides further support to distinguish the two different heat stress responses (HSR and HTR).

As mentioned earlier, the process of regulating both responses require highly coordinated gene expression. In order to accomplish this, they must be able to sense and respond rapidly to the heat stimuli. How can this regulation be accomplished? The next section will discuss the roles of RNA as regulators of gene expression.

2.4 Overview of regulatory RNAs in bacteria

Bacteria are very versatile and adaptive microorganisms. In order to cope with the dynamic environment they live in, different groups of regulator RNAs are used to adjust their physiological state.

Riboswitches are one group of regulatory RNA that modulates gene expression in response to small molecules in the environment. They are part of the mRNA that they regulate. The leader sequence of a riboswitch folds into structures that change when small molecules are bound to them. This enables them to sense nutrient availability around the cell (Grundy and Henkin, 2006). Extensive review for riboswitches have been published (Coppins *et al.*, 2007; Winkler and Breaker, 2005)

Another class of regulatory RNA is the recently discovered CRISPR (clustered regularly interspaced short palindromic repeats). These RNAs confer resistance to bacteriophage infection (Sorek *et al.*, 2008) and interfere in plasmid conjugation (Marraffini and Sontheimer, 2008). CRISPR sequences are highly variable DNA regions with homology to bacteriophage and plasmid sequences. They consist of an approximately 550 bp leader sequence and a series of repeat-spacer units downstream of the leader sequence. About 40% of CRISPR sequences have been found in bacteria while as much as 90% have been detected in archaea, thus emphasizing its potential broad range importance (Sorek *et al.*, 2008).

Finally, one of the most extensively studied subset of regulatory RNAs is the small RNA (sRNA). This group of regulators exerts their effect most commonly by base pairing with RNAs. There are also sRNAs that modulate gene expression by interacting directly with proteins (protein-binding sRNA). The next two sections will discuss the mechanisms and functions of sRNAs in detail.

2.5 Characteristics of sRNA

The first bacterial sRNA, RNAI, was discovered in 1981. RNAI is 108 nucleotides long and was found to block ColE1 plasmid replication in *E. coli* by the base-pairing mechanism (Stougaard *et al.*, 1981; Tomizawa *et al.*, 1981). The first

chromosomally-encoded sRNA was discovered in 1984. It is MicF RNA (174 nucleotides) from *E. coli*. This sRNA blocks the translation of the *ompF* gene that encodes the outer membrane porin (Mizuno *et al.*, 1984). Both of these sRNAs and many other early sRNAs were serendipitously discovered due to their abundance in the cell (Wassarman *et al.*, 1999) or by accident when studying other things. A few examples of sRNAs discovered in this manner are MicF, DsrA, and CsrB.

Small RNAs have been found to exist in both eukaryotes (Brodersen and Voinnet, 2009) and prokaryotes (Waters and Storz, 2009). Although small RNAs play similar roles in modulating gene expression in both types of cells, the nomenclature of this class of regulatory RNA has remained non-uniform (Storz and Haas, 2007). In eukaryotes, the term non-coding RNA (ncRNA) has been the predominant form of naming small RNA although there are papers (Gottesman, 2005; Toledo-Arana *et al.*, 2007; Repoila and Darfeuille, 2009) that used this term to describe bacterial small RNA. MicroRNA (miRNA) is also used widely to denote eukaryotic small RNA. Another common term used to describe bacterial regulatory RNA is "sRNA". However, there are bacterial RNAs that act as regulators, are capable of encoding proteins and are also not very small such as RNAIII (514 nt) in *Staphylococcus aureus* (Novick and Geisinger, 2008). Therefore, by definition it is neither non-coding nor small in size. Nevertheless, since the term sRNA has been widely used in recent bacterial literature (Storz and Haas, 2007), this term will be used in this thesis to denote regulatory small RNAs in bacteria.

The length of sRNAs is equally ambiguous with many different authors giving variable size ranges in their publications (Frohlich and Vogel, 2009; Gottesman, 2005; Majdalani *et al.*, 2005; Liu *et al.*, 2009). However, it can be summarized as a broad range between 20 nt to 550 nt.

As of 2009, no three-dimensional structures have been solved for bacterial sRNAs (Waters and Storz, 2009). sRNAs that have been experimentally probed for secondary structure are also limited. The general sRNA secondary structure begins with sequences that are able to fold into a stem-loop and end with a rho-independent transcription terminator. These structures help to stabilize the sRNA and it has been shown via *in vivo* stability tests that sRNAs are significantly more stable than mRNAs (Vogel *et al.*, 2003; Masse *et al.*, 2003).

sRNA, just like any other RNA, has to be synthesized in order for it to act. The bacterial sRNA transcripts are generally not processed as they do not undergo further post-transcriptional modification after being synthesized. These highly structured sRNAs usually use the same promoter as other bacterial genes and are ready to be used upon synthesis (Gottesman, 2005). Synthesis of miRNA in eukaryotes, on the contrary, requires preprocessing and export into the cytoplasm. It is then further processed by a ribonuclease known as the Dicer ribonuclease to generate the active form (approximately 22 nt) before it is functional in the cells. In some cases, multiple miRNA are produced by a single transcript (Xu *et al.*, 2004). sRNA is expressed as part of the well-known stress response systems such as iron limitation, oxidative stress, low temperature, and carbon metabolism. sRNAs are also synthesized to modulate virulence and pathogenesis, biofilm formation, outer membrane synthesis, and quorum sensing. Due to its wide array of regulatory functions, it was postulated that all major stress responses will involve at least one sRNA as part of its regulatory pathway (Gottesman, 2005). Figure 2.1 summarizes the involvement of sRNAs in some of the major stress responses in bacteria.

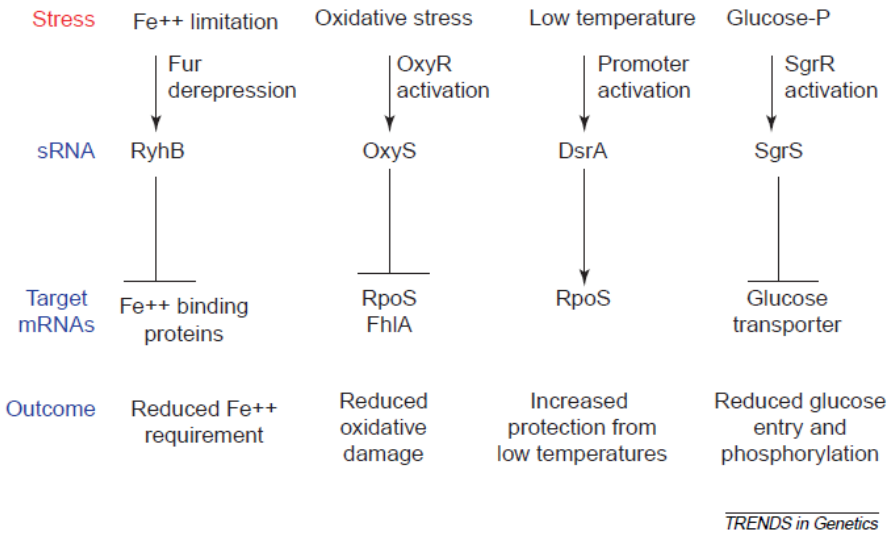

Figure 2.1 sRNAs involved in some of the major stress responses in bacteria. Figure from (Gottesman, 2005) (Figure 2, Page 401, GOTTESMAN, S. 2005. Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends Genet,* 21**,** 399-404)

2.6 Regulatory mechanisms of sRNA

2.6.1 Levels of gene regulation by sRNA

There are two major regulatory mechanisms that are used by sRNAs to modulate gene expression in cells. They can either base-pair to target messenger RNA (mRNA), forming an RNA duplex and affecting the fate of the mRNA, or bind to proteins and modify their activity. sRNAs can exert their effect at the post-transcriptional, translation, and post-translational level. The sRNA-mRNA pairing can result in either gene expression activation (transcription initiation, translation initiation) or gene expression repression (mRNA degradation, transcription termination, translation blockage). Figure 2.2A (*cis*) and 2.2B (*trans*) illustrates the level of gene regulation by sRNAs.

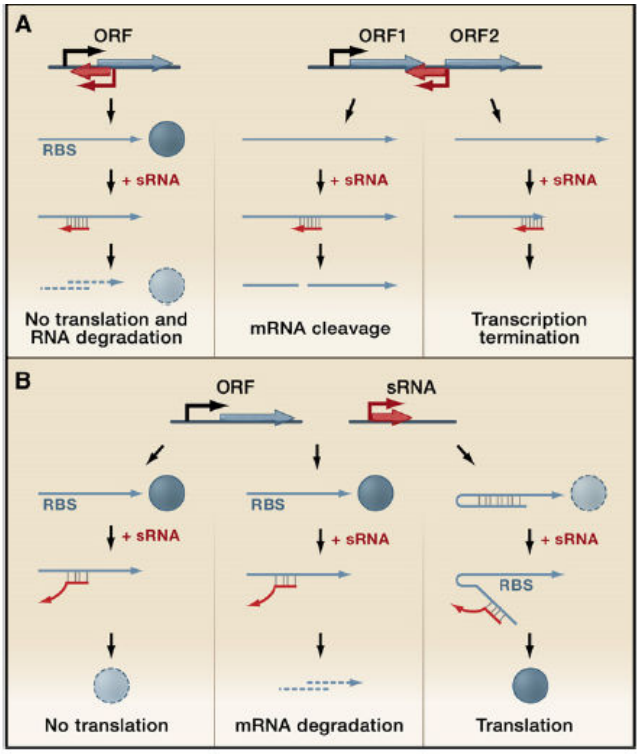

Figure 2.2 Regulation mechanism by sRNA via base-pairing with target mRNA. (A) Regulation by *cis*-encoded sRNA. (B) Regulation by *trans*-encoded sRNA. RBS: Ribosomal binding site. Figure from Waters *et al.,* 2009 (Figure 2, Page 618, WATERS, L. S. & STORZ, G. 2009. Regulatory RNAs in bacteria. *Cell,* 136**,** 615-28).

2.6.2 sRNA regulation via base-pairing mechanism

Base-pairing is the most common form of regulation by sRNA (Storz *et al.*, 2005). sRNAs in this group can further be separated into two classes based on the location they are found. The *cis*-encoded sRNAs have extensive base-pairing (at least 75 nt) with its target mRNA (Brantl, 2007; Wagner *et al.*, 2002), while the *trans*-encoded sRNAs have limited complimentarity with its target and may require the help of an RNA chaperone, Hfq, for its action (Aiba, 2007; Gottesman, 2005).

2.6.2.1 *cis*-encoded sRNA

cis-encoded sRNAs are encoded on the same gene locus as the target mRNA but is transcribed in the opposite direction as discrete RNA species. This enables it to maintain a perfect complementarity to their targets. This RNA species is diffusible and highly structured with one to four stem-loops (Brantl, 2007). In the eukaryotic system, small interfering RNA (siRNA) is analogous to this class of bacterial sRNA (Storz *et al.*, 2005). These sRNAs are approximately 21-25 nucleotides in length, derived from exogenous double-stranded RNA (Meister and Tuschl, 2004). They have been proposed to confer resistance against foreign RNA (Montgomery, 2004).

The majority of *cis*-encoded sRNAs are located at mobile genetic elements such as plasmids, transposons, and bacteriophages (Brantl, 2002). This enables the sRNA to maintain the appropriate copy number of mobile genetic elements via inhibition of primer replication or transposase translation (Brantl, 2007). Plasmid encoded sRNAs are expressed constitutively while chromosomally-encoded *cis*-encoded sRNAs are expressed only under specific conditions. For example, GadY is expressed during stationary phase in *E. coli* (Opdyke *et al.*, 2004) and IsrR in *Synechocystis* during iron depletion conditions (Duhring *et al.*, 2006). Chromosomal

versions of this sRNA are also being discovered but their physiological roles are still less understood (Waters and Storz, 2009; Brantl, 2007).

Another group of *cis*-encoded sRNAs acts as an antitoxin, repressing the translation of toxic proteins that kill when they lose their mobile elements. In *E. coli* the *hok/sok* loci are present in plasmids (plasmid R1 and F plasmid) (Gerdes *et al.*, 1990) and the chromosome (Pedersen and Gerdes, 1999). The expression of *hok* (host killing) mRNA is suppressed by Sok RNA (suppressor of killer). The loss of plasmid expressing the Sok RNA has been shown to kill the cells, but the effect of losing chromosomally-encoded sRNA remains a debatable subject. One proposed model of chromosomally-encoded sRNA is that the toxin-antitoxin module promotes slow growth of cells that are under stress, giving them time to recover and adjust to the new environment (Unoson and Wagner, 2008; Kawano *et al.*, 2007). Another proposed model is that the chromosomal copy serves as a defense against plasmids that are homologous to it, inhibiting the expression of the toxin encoded by the plasmid.

There are also *cis*-encoded sRNAs that control the expression of genes within an operon. They are encoded at sequences between two ORFs with complementary sequences. In *E. coli*, GadY sRNA (stationary phase induced) is encoded between the *gadX* and *gadW* genes. GadY sRNA has complementary sequence to *gadXW* mRNA and causes the cleavage of *gadXW* to *gadX* and *gadW*, increasing *gadW* levels in the cell during stationary phase (Opdyke *et al.*, 2004; Tramonti *et al.*, 2008).

2.6.2.2 *trans*-encoded sRNA

trans-encoded sRNAs are encoded at distinct chromosomal locations relative to their target mRNA. As described earlier, they only have partial complementarity with their targets. The function of many chromosomal *trans*-encoded sRNAs have been elucidated. The function of micro RNA (miRNA) in eukaryotes has been said to be analogous to bacterial *trans*-encoded sRNA (Waters and Storz, 2009; Aiba, 2007; Gottesman, 2005). They are approximately 22 nucleotides in length, being cleaved from longer stem-loop RNAs by the same enzymes that produce siRNA (*cis*-encoded RNA) (He and Hannon, 2004; Bartel, 2004; Nelson *et al.*, 2003; Storz *et al.*, 2005).

The locations of the target mRNA and *trans*-encoded sRNA have very little correlation. Each *trans*-encoded sRNA is capable of base-pairing with multiple mRNAs (Papenfort and Vogel, 2009). This capability can be explained by the fact that *trans*-encoded sRNAs require a limited amount of direct contact with the target mRNA. In contrast to *cis*-encoded sRNAs which require perfect complementation with the target mRNA, *trans*-encoded sRNAs use a discontinuous patch of complementary sequence to exert its effect. As little as 10 to 25 nucleotides are required for successful interaction with target mRNA, but much less nucleotides are actually used in regulating the target mRNA. In *E. coli*'s SgrS sRNA, which represses *ptsG* in glucose metabolism, it has a 23 nucleotide base-pair potential with *ptsG*. However, only 4 single nucleotide mutations in SgrS are significant in the downregulation of *ptsG* (Kawamoto *et al.*, 2006). The fact that a single sRNA can modulate a global response for a specific physiological response makes it similar to transcription factor but acts at the posttranscriptional level (Valentin-Hansen *et al.*, 2007; Masse *et al.*, 2007; Bejerano-Sagie and Xavier, 2007).

The majority of *trans*-encoded sRNAs are transiently synthesized under very specific growth phase or environmental conditions (Repoila and Darfeuille, 2009). This is in contrast to *cis*-encoded sRNAs that are mostly expressed constitutively. The involvement of *trans*-encoded sRNAs in specific environmental conditions is discussed in section 2.6.4.1.

2.6.3 Facultative requirement of Hfq protein (RNA chaperone)

Hfq (host factor required for phage Qβ RNA replication) protein is a RNA chaperone that facilitates the base-pairing between sRNAs and their mRNA targets. This hexameric protein is often required in *trans*-encoded sRNA regulation of mRNA to facilitate the limited complementarity between the sRNA and target mRNA (Aiba, 2007; Brennan and Link, 2007; Valentin-Hansen *et al.*, 2004). Hfq belongs to the Sm proteins family in eukaryotes which are used in splicing and mRNA decay (Sun *et al.*, 2002; Schumacher *et al.*, 2002). It functions by remodeling RNAs bound to it and melting secondary structures.

Almost 50% of the sequenced bacterial genomes contain Hfq homologs (Sun *et al.*, 2002). Although many bacteria have the homolog of this protein, the requirement of Hfq for sRNA regulation is not compulsory. Jousselin and collegues proposed that the need of Hfq is dependent on the overall GC content of bacterial genomes, the free energy for sRNA-mRNA pairing, the genome size, and the structural variation among Hfq proteins (Jousselin *et al.*, 2009).

E. coli requires Hfq protein for its *trans*-encoded sRNA-mRNA interaction. A minimum of 22 sRNAs have been reported to require Hfq to function (Majdalani *et al.*, 2005). In a *hfq* deletion strain, sRNAs that require Hfq have a shorter half-life compared to wild type (Urban and Vogel, 2007). RNase E is also linked to Hfq. It is

recruited for rapid target mRNA degradation or translation repression but the sRNA is protected from degradation (Morita *et al.*, 2005). This could probably be due to the fact that Hfq binding sites on the sRNA is shared by the RNase E cleavage sites (Moll *et al.*, 2003). Hence, Hfq might shelter the sRNA from endonucleolytic attack and increases its half-life (Urban and Vogel, 2007). Recently it has been reported that the requirement for Hfq can be bypassed in *rpoS* mRNA regulation by sRNA in *E. coli* if the stability of the sRNA-mRNA complex is high (Soper *et al.*, 2010). This is contrary to a previous study (Sledjeski *et al.*, 2001) that shows that Hfq is a must for *rpoS* mRNA regulation.

In the pathogens *Vibrio cholerae*, *Vibrio harveyi*, and *Listeria monocytogenes*, Hfq requirement is facultative. Four sRNAs in *V. cholerae* (Qrr1-Qrr4), which affects gene expression during high cell density (Lenz *et al.*, 2004), requires Hfq to function (Hammer and Bassler, 2007). However, *V. harveyi* VrrA, which downregulates *ompA* porin mRNA, functions in the absence of Hfq *in vivo* (Song *et al.*, 2008). In *L. monocytogenes*, co-immunoprecipitation with Hfq reveals three sRNAs (LhrA-LhrC), indicating that Hfq is required for this sRNA to function (Christiansen *et al.*, 2006). However, there are sRNAs (RliB, RliE, RliI) from this bacterium that regulate their targets *in vivo* without the need of Hfq (Mandin *et al.*, 2007).

In some Gram-positive bacteria, Hfq requirement by sRNA-mediated gene regulation is facultative. Relative to the abundance of Hfq protein in *E. coli*, the *hfq* gene is present at very low level in *S. aureus*, a low GC Gram-positive bacterium (Bohn *et al.*, 2007). The *hfq* deletion mutant did not show any significant phenotypic differences when cultured at 1000 different conditions as compared to wild type. In *B. subtilis*, an Hfq homolog is encoded by the *ymaH* gene (Silvaggi *et al.*, 2005).

Two experimentally studied sRNAs, SR1 (Heidrich *et al.*, 2006) and FsrA (Gaballa *et al.*, 2008), do not require YmaH for their function. These findings suggest that Hfq is not compulsory in sRNA-mRNA regulation (Bohn *et al.*, 2007).

2.6.4 Positive and negative regulation by sRNA base-pairing

2.6.4.1 Negative regulation

trans-encoded sRNA gene regulation mostly results in downregulation of its target mRNA. This negative regulation could be achieved by translational inhibition, mRNA degradation, or a combination of both mechanisms (Waters and Storz, 2009; Aiba, 2007; Gottesman, 2005).

Translational inhibition is achieved by the binding of sRNA to the 5' untranslated region (5' UTR) of the target mRNA, which blocks the binding of ribosome to the ribosome binding site (RBS). There are sRNAs that bind far upstream from the start codon but manage to repress the translation of its target mRNA, such as GcvB and RyhB sRNA.

2.6.4.2 Positive regulation

Although most of the sRNA-mRNA base-pairing involves negative regulation, there are 2 direct base-pairing mechanisms known to date that enable sRNA to positively regulate its target mRNA. They are the "anti-antisense mechanism" and 3' transcript stabilization (Frohlich and Vogel, 2009). There is also an indirect regulation mechanism that uses RNA mimicry, such as GlmYZ sRNA (described in section 2.6.5) and MicM sRNA, that activates the gene expression (Rasmussen *et al.*, 2009). Figure 2.3 summarizes all of these mechanisms.